VEGETATION CLASSIFICATION AND SURVEY

International Association for Vegetation Science (IAVS)

Vegetation Classification and Survey 2: 59–63
doi: 10.3897/VCS/2021/64911

LONG DATABASE REPORT ECOINFORMATICS

SWEA-Dataveg: A vegetation database for sub-Saharan Africa

Miguel Alvarez[1], Michael Curran[2], Itambo Malombe[3]

1 *Institute of Crop Science and Resource Conservation, University of Bonn, Bonn, Germany*
2 *Department of Socio-Economic Sciences, FiBL, Frick, Switzerland*
3 *East African Herbarium, National Museums of Kenya, Nairobi, Kenya*

Corresponding author: Miguel Alvarez (malvarez@uni-bonn.de)

Academic editor: Florian Jansen ♦ **Received** 24 February 2021 ♦ **Accepted** 28 March 2021 ♦ **Published** 5 May 2021

Abstract

SWEA-Dataveg is a vegetation-plot database collecting observations mainly in sub-Saharan Africa but also open to the rest of the African continent. To date this database contains more than 5,500 plot observations provided by 47 sources (projects, monographs, and articles). While the database is stored in PostgreSQL (including the PostGIS extension), the R-package "vegtable" implements a suitable exchange format. In this article we assess the current content of SWEA-Database and introduce its history and future as a repository of data for syntaxonomic assessments and macroecological research.

Keywords

ecoinformatics, plant biodiversity, taxlist, syntaxonomy, vegetation ecology, vegtable

Introduction

In sub-Saharan Africa as elsewhere, documenting and classifying vegetation has become an urgent task to enable the proper assessment of endangered ecosystems (Jansen et al. 2016). With an increasing number of research projects dealing with vegetation ecology in the region, there is a vast amount of information of high scientific value that could be made accessible to the wider research community. At the same time, knowledge accumulated in past research programs can also provide the basis for constructive research into vegetation history, biogeography and conservation. Database structures such as vegetation-plot databases may serve as important repositories for data curation and ensure research repeatability and meta-analysis in the context of macroecological and biogeographical studies (Dengler et al. 2011; Bruelheide et al. 2019).

The database SWEA-Dataveg (Alvarez et al. 2012b) was initiated as a repository for ongoing projects in East Africa, specifically for the SWEA project (Agricultural use and vulnerability of small wetlands in East Africa). At its genesis the database was focusing on the collection of data from wetland ecosystems in Kenya and Tanzania (see Alvarez et al. 2012a). Through follow-up projects and collaboration activities with the ETH-Zürich (Switzerland) and the East African Herbarium (Kenya), the database was expanded to all vegetation formations and included data from additional African countries.

This report briefly displays the current status of the vegetation-plot database SWEA-Dataveg (GIVD AF-00-006) and its applications in the research of vegetation ecology and biogeography in sub-Saharan Africa.

PENSOFT

Copyright *Miguel Alvarez et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

GIVD Fact Sheet

GIVD Database ID: AF-00-006			**Last update:** 2021-02-23
SWEA-Dataveg		**Web address:** https://kamapu.github.io/sweadataveg.html	
Database manager(s): Miguel Alvarez (kamapu78@gmail.com)			
Owner: Miguel Alvarez (as custodian)			
Scope: Relevés in small wetlands of Kenya and Tanzania collected during the sampling activities of the SWEA project, including semi-natural vegetation (non-used or light used fields), fallows, grasslands and weed communities in crops. Currently the GlobE wetlands project is continuing data collection with a similar scope as SWEA. Additional information from other projects and published relevés from East Africa are considered. Though this database is currently not freely available, a delivery for free use after end of GlobE wetlands project is considered.			
Abstract: SWEA (agricultural use and vulnerability of small wetlands in East Africa) is a multidisciplinary project whose task is to evaluate the effects of land use on the ecological and socio-economical functions of small wetlands in Kenya and Tanzania. In order to allow the availability of the collected data for further studies we stored them into SWEA-Dataveg, a database performed in Microsoft Access (mdb-format). Because this project is dealing not only with vegetation science but also with geography, soil science, hydrology and socio-economy, the database also contains information related with these research fields. Additionally some functional traits of the plant species occurring in the relevés are included in the species list. The sampling areas are concentrated in four localities, two of them in Kenya (Karatina and Rumuruti) and two in Tanzania (Malinda and Lukozi). The vegetation ecology group is dealing in the project with the classification of the vegetation according with the species composition, the correlation of plant communities with environmental factors and land uses, and the survey of potential indicator species for the detriment on the resilience of wetlands. Once finished the storage we are considering an adaptation of SWEA-Dataveg into a Turboveg-format as well as its extension to further projects (e.g. SWEA phase II) and relevés collected from publications.			
Availability: according to a specific agreement		**Online upload:** no	**Online search:** no
Database format(s): PostgreSQL		**Export format(s):** other, vegtable (R)	
Plot type(s): normal plots		**Plot-size range:** 0.5 to 10000	
Non-overlapping plots: 5552	**Estimate of existing plots:** 10000	**Completeness:** 56%	**Status:** completed and continuing
Total no. of plot observations: 5552	**Number of sources (biblioreferences, data collectors):** 33		**Valid taxa:** 3403
Countries (%): KE: 58; TZ: 10; BJ: 1; CD: 10; ET: 4; RW: 3; TG: 3; UG: 9; ZM: 3			
Formations:			
Guilds: all vascular plants: 100%			
Environmental data (%):			
Performance measure(s): presence/absence only: 1%; cover: 99%			
Geographic localisation: point coordinates less precise than GPS, up to 1 km: 100%			
Sampling periods: 1930-1939: 5%; 1940-1949: 2%; 1950-1959: 1%; 1960-1969: 1%; 1980-1989: 17%; 1990-1999: 14%; 2000-2009: 7%; 2010-2019: 26%; unknown: 27%			
Information as of 2021-02-23 further details and future updates available from http://www.givd.info/ID/AF-00-006			

History

The idea of establishing a vegetation-plot database started during a visit to the 8th Meeting on Vegetation Databases, held at the University of Greifswald, Germany, in 2009. The project was officially launched in 2010 and the first report was published in 2012 with a small collection of 206 plots originally stored in a Microsoft-Access database (Alvarez et al. 2012b). Since then, this database has been affiliated with research activities at the University of Bonn, Germany, in collaboration with diverse academic and research institutions in Eastern Africa.

In 2015, and in the context of a collaborative activity between the SWEA-Project and the ETH-Zürich, Switzerland, SWEA-Dataveg migrated to the software Turboveg (Hennekens and Schaminée 2001) and the first trials for data exchange and processing using R-images and R-scripts were carried out. At that time, export of Turboveg to R was completed using the package "vegdata" (Jansen and Dengler 2010).

After the first releases of the packages "taxlist" and "vegtable" at CRAN in 2017 (see Alvarez and Luebert 2018), the database migrated again, this time to PostgreSQL including the PostGIS extension for handling the location of plots in a Geographical Information System (GIS). During this development, SWEA-Database became larger and more complex, and partially interlinked with the database "sudamerica" (former CL-Dataveg, GIVD SA-CL-001; Alvarez et al. 2012c).

Content of the database

Currently, the database contains 5,552 plot observations (relevés) collected from 47 sources, including projects, journal articles and monographs. These observations contain records of 3,530 plant species belonging to 1,318 genera and 216 families. The dominant families are *Leguminosae* (402 species; 10.4%), *Poaceae* (393 species; 10.2%), *Compositae* (290 species; 7.5%), and *Cyperaceae* (212 species; 5.5%).

According to record date and year of publication, the oldest observations are from 1937 (Lebrun 1947, 1960), while the most recent records are from 2020 (unpublished data). Plot sizes comprise < 1 m^2 (37 plots, 0.7%); 1–10 m^2 (1,168 plots, 21.0%); 10–100 m^2 (1,289 plots, 23.2%); 100–1,000 m^2 (616 plots, 11.1%); 1,000–10,000 m^2 (84 plots, 1.5%); and for 2,358 observations (42.5%) the plot size is unknown. A total of 1,822 plot observations (32.8%) were collected in projects affiliated to SWEA-Dataveg.

The current version of SWEA-Dataveg is stored in a PostgreSQL database, including the PostGIS extension for geo-referenced information. Plot observations are organ-

ized in a table called "header" and linked to several tables analogous to the popup tables of Turboveg (Hennekens and Schaminée 2001). A taxonomic list is also integrated into this database, following the structure used by the R-package "taxlist" (Alvarez and Luebert 2018). Data export is preferentially designed in SQL language and assigned to a "vegtable" object in R (see https://github.com/kamapu/vegtable). Further process and assessment can be done either in R or exporting to any spreadsheet application for analysis. Additionally, export to the software Juice (Tichý 2002) is carried out by a function called "write_juice()".

All plots included in the database are geo-referenced. A logical variable called "validation_coordinates" indicates whether these coordinates were provided by the authors as coordinate values or in a detailed map ("true"), or if they are inferred from the description of locality ("false"). Observations have been undertaken in 12 countries with 2,804 plots (51%) sampled in Kenya, 986 (18%) in the Democratic Republic of the Congo, 467 (8%) in Ethiopia, and 425 (8%) in Tanzania. The rest of the plots were collected in Uganda, Togo, Rwanda, South Africa, Burundi, Congo-Brazzaville, Benin, and Zambia (see Figure 1).

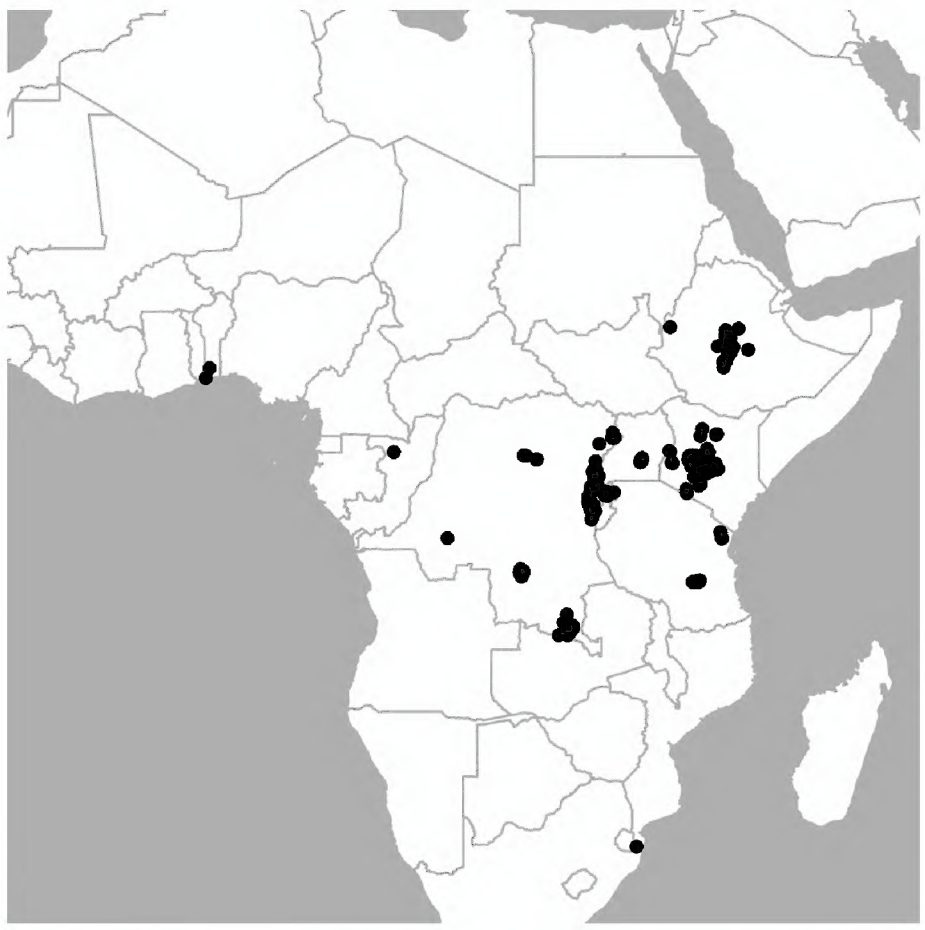

Figure 1. Geographical distribution of plot observations (black dots) stored in SWEA-Dataveg.

SWEA-Dataveg attempts to collect as much of the information originally published with plot observations as possible. Besides information on plot size, recording dates and locations (coordinates and descriptions of localities), additional data on slope inclination, exposition, elevation, total vegetation cover, soil physical and chemical properties and remarks, if provided by the sources, are digitized and stored. From all observations, 79% are stored with a sampling date, 64% with coordinates, 58% with information on plot size, and 21% with information on soil physical or chemical properties (Figure 2). Furthermore, original pages and table number as well as assignment to a specific plant community is also documented.

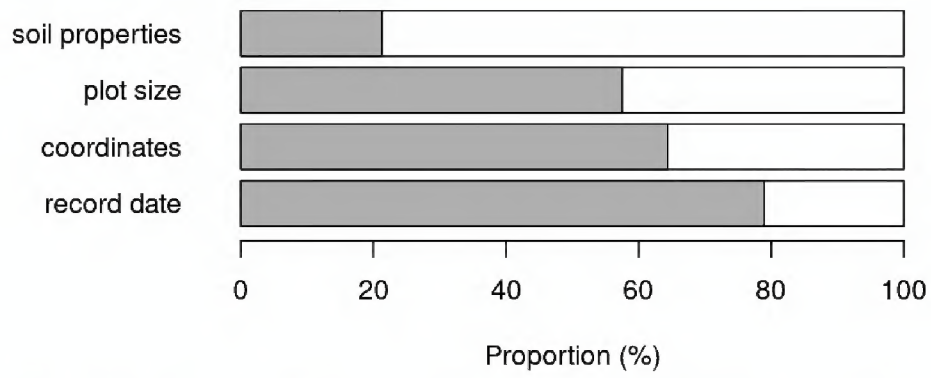

Figure 2. Completeness of important information within the plots stored in SWEA-Dataveg. Grey areas represent the proportion of observations containing any data for the respective variables.

The associated taxonomic list is supported by five sources referred to as taxon views (see Alvarez and Luebert 2018). This module contains information on taxonomic ranks, parent-child relationships (e.g. indication of the parent genus for a species) and taxon attributes (e.g. life forms, chorology and functional traits). The later information is usually collected from secondary references, including on-line databases, and complements specific project objectives.

Additional features

All data sources are supported by a private soft copy of the relevant published article to enable cross-validation of fidelity of data stored in the database. Digitization procedures strive to resemble the data published in the original source.

Projects attempting to derive critical assessments of classifications in the context of the Braun-Blanquet approach (e.g. Alvarez 2017) are also catered for with a collection of syntaxonomic nomenclatures and Cocktail algorithms stored as "expert systems" (see Landucci et al. 2015).

Besides all of these features, the development of the R-packages "taxlist" and "vegtable" (Alvarez and Luebert 2018) are strongly dependent on the assessment of data contained in SWEA-Dataveg and are used as the main mechanisms for data sharing and publication. The implementation of R-scripts in the assessment of data assure the repeatability of statistics while the current efforts to integrate r-markdown in some functions enables the possibility of producing automatic updates of summaries such as lists of data per syntaxa and publications or check lists of plant species.

At present, data is accessible only after special agreements with the custodian. While data stored from ongoing projects are highly restricted at least during the lifespan of the respective projects, we expect to be able to make data freely available from already published works. The preferred format for exchange is an R-Image including a vegtable object (Alvarez and Luebert 2018). Further alternative formats are Juice tables, SQL dump files for freeware relational database systems (e.g. PostgreSQL, MySQL, LibreOffice Base), and spreadsheets in xlsx, odt and csv formats. In all of these cases, the content of the requested files requires correspondence with the custodian.

Resulting publications

From its origins, SWEA-Dataveg focused on a preliminary classification of wetland vegetation in East Africa (Alvarez et al. 2012a). This work was followed by a classification of aquatic and semi-aquatic vegetation using observations collected in 2012 in Kenya and Tanzania and addressing the Braun-Blanquet approach (Alvarez 2017).

In the specific case of Kenya, a model describing plant biodiversity and spatial conservation prioritization was performed for the Kenyan subset and included a pool of bioclimatic, macroecological and economic factors as explanatory variables (Scherer et al. 2017a, b). This work inferred locations in the country that are most suitable for the expansion of protected areas in order to meet national targets for biodiversity conservation and estimated the required funding to achieve this.

SWEA-Dataveg also supported the design of ecological assessment and monitoring methods, such as an adaptation of the WET-Health approach by Beuel et al. (2016), and the use of physiognomic properties of the vegetation for the estimation of the biological integrity completed by Behn et al. (2018). In both works, the information included in the database was used for the calibration of regression models and the evaluation of outcomes.

Ongoing projects are dealing with distribution models of invasive species in Eastern Africa, in particular on *Prosopis juliflora* (Sw.) DC. (Alvarez et al. 2019) and *Parthenium hysterophorus* L. (no publications to date).

In addition to inclusion in the Global Index of Vegetation-Plot Databases (Dengler et al. 2011; Alvarez et al. 2012b), this database also contributed to the sPlot initiative (Bruelheide et al. 2019).

The way forward

The implementation of a multiple-taxon views approach, for instance considering discrepancies among different projects involved in the African Plant Database (https://www.ville-ge.ch/musinfo/bd/cjb/africa/recherche.php) and some regional floras (e.g. Flora of Tropical East Africa, Beentje et al. 1952–2012; Flora of Ethiopia and Erithrea, Hedberg et al. 1989–2009), will make this database more versatile. This will also allow it to expand areas of coverage and to integrate other databases under the same database model, such as the database "sudamerica" (Alvarez et al. 2012c).

We also seek to integrate an electronic document library, which is at present housed in a separated database formatted as a BibTeX file and linked to respective data sources as well as taxonomic and syntaxonomic authorities.

Acknowledgements

The database is currently maintained in the context of the project "Future Invasions" within the Collaborative Research Centre "Future Rural Africa" (http://www.futureruralafrica.de/). We thank Mrs. Emilia Lösche for her support accessing the valuable collections of the Library of the Geographical Department at the University of Bonn in Germany. Several students have supported the work of digitizing data and testing assessments by the developed R-packages, to whom we are very thankful.

References

Alvarez M (2017) Classification of aquatic and semi-aquatic vegetation in two East African sites: Cocktail definitions and syntaxonomy. Phytocoenologia 47: 345–364. https://doi.org/10.1127/phyto/2017/0078

Alvarez M, Luebert F (2018) The taxlist package: Managing plant taxonomic lists in R. Biodiversity Data Journal 6: e23635. https://doi.org/10.3897/BDJ.6.e23635

Alvarez M, Becker M, Böhme B, Handa C, Josko M, Kamiri HW, Langensiepen M, Menz G, Misana S, ..., Sakané N (2012a) Floristic classification of the vegetation in small wetlands of Kenya and Tanzania. Biodiversity & Ecology 4: 63–76. https://doi.org/10.7809/b-e.00060

Alvarez M, Möseler BM, Josko M, Becker M, Langensiepen M, Menz G, Böhme B, Oyieke HA, Handa C, ..., Sakané N (2012b) SWEA-Dataveg – vegetation of small wetlands in East Africa. Biodiversity & Ecology 4: 294–295. https://doi.org/10.7809/b-e.00092

Alvarez M, Möseler BM, Martín CS, Ramírez C, Amigo J (2012c) CL-Dataveg – a database of Chilean grassland vegetation. Biodiversity & Ecology 4: e443. https://doi.org/10.7809/b-e.00230

Alvarez M, Heller G, Malombe I, Matheka KW, Choge S, Becker M (2019) Classification of *Prosopis juliflora* invasion in the Lake Baringo basin and environmental correlations. African Journal of Ecology 57: 296–303. https://doi.org/10.1111/aje.12601

Beentje HJ [Ed.] (1952–2012) Flora of Tropical East Africa. 225 vols. Kew Botanical Gardens, London.

Behn K, Becker M, Burghof S, Möseler BM, Willy DK, Alvarez M (2018) Using vegetation attributes to rapidly assess degradation of East African wetlands. Ecological Indicators 89: 250–259. https://doi.org/10.1016/j.ecolind.2018.02.017

Beuel S, Alvarez M, Amler E, Behn K, Kotze DC, Kreye C, Leemhuis C, Wagner K, Willy DK, ..., Becker M (2016) A rapid assessment of anthropogenic disturbances in East African wetlands. Ecological Indicators 67: 684–692. https://doi.org/10.1016/j.ecolind.2016.03.034

Bruelheide H, Dengler J, Jiménez-Alfaro B, Purschke O, Hennekens SM, Chytrý M, Pillar VD, Jansen F, Kattge J, ..., Zverev A (2019) sPlot – a new tool for global vegetation analyses. Journal of Vegetation Science 30: 161–186. https://doi.org/10.1111/jvs.12710

Dengler J, Jansen F, Glöckler F, Peet RK, De Cáceres M, Chytrý M, Ewald J, Oldeland J, Lopez-Gonzalez G, ..., Spencer N (2011) The global index of vegetation-plot databases (GIVD): A new resource for vegetation science. Journal of Vegetation Science 22: 582–597. https://doi.org/10.1111/j.1654-1103.2011.01265.x

Hedberg I et al. [Eds] (1989–2009) Flora of Ethiopia and Eritrea. 10 vols. National Herbarium, Addis Ababa University, Addis Abeba, ET.

Hennekens SM, Schaminée JHJ (2001) TURBOVEG, a comprehensive data base management system for vegetation data. Journal of Vegetation Science 12: 589–591. https://doi.org/10.2307/3237010

Jansen F, Dengler J (2010) Plant names in vegetation databases – a neglected source of bias. Journal of Vegetation Science 21: 1179–1186. https://doi.org/10.1111/j.1654-1103.2010.01209.x

Jansen F, Bergmeier E, Dengler J, Janišová M, Krestov P, Willner W (2016) Vegetation classification: A task of our time. Phytocoenologia 46: 1–4. https://doi.org/10.1127/phyto/2016/0134

Landucci F, Tichý L, Šumberová K, Chytrý M (2015) Formalized classification of species-poor vegetation: A proposal of a consistent protocol for aquatic vegetation. Journal of Vegetation Science 26: 791–803. https://doi.org/10.1111/jvs.12277

Lebrun J (1947) Exploration du Parc National Albert. Mission J. Lebrun (1937–1938). 1. La végétation de la plaine alluviale au sud du Lac Édouard. Institut des Parcs Nationaux du Congo Belge, Bruxelles.

Lebrun J (1960) Exploration du Parc National Albert. Mission J. Lebrun (1937–1938). 2. Études sur la flore et la végétation des champs de lave au nord du Lac Kivu (Congo Belge). Institut des Parcs Nationaux du Congo Belge, Bruxelles, 352 pp.

Scherer L, Curran M, Alvarez M (2017a) Expanding Kenya's protected areas under the convention on biological diversity to maximize coverage of plant diversity. Conservation Biology 31: 302–310. https://doi.org/10.1111/cobi.12792

Scherer L, Curran M, Alvarez M (2017b) Staggering financial shortfalls to meet biodiversity targets in Kenya. Atlas of Science February. http://atlasofscience.org/staggering-financial-shortfalls-to-meet-biodiversity-targets-in-kenya

Tichý L (2002) JUICE, software for vegetation classification. Journal of Vegetation Science 13: 451–453. https://doi.org/10.1111/j.1654-1103.2002.tb02069.x

E-mail and ORCID

Miguel Alvarez (Corresponding author, malvarez@uni-bonn.de), ORCID: 0000-0003-1500-1834
Michael Curran (michael.curran@fibl.org), ORCID: 0000-0002-1858-5612
Itambo Malombe (imalombe@museums.or.ke), ORCID: 0000-0002-8168-7013